UNITED NATIONS

ECONOMIC
AND
SOCIAL COUNCIL

Distr.
LIMITED

E/ECA/PSD.3/36
3 November 1983

Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA

Third session of the Joint Conference
of the African Planners, Statisticians
and Demographers

Addis Ababa, 5 - 14 March 1984

DEMOGRAPHIC DATA COLLECTION AND ANALYSIS IN THE AFRICAN REGION - AN APPRAISAL OF THE QUALITY OF DATA

TABLE OF CONTENTS — PARAGRAPHS

1. Introduction 1 - 7
2. Types and varieties of data collected 8 - 12
3. Quality of Data 13 - 41
4. Conclusions and Recommendations 42

DEMOGRAPHIC DATA COLLECTION AND ANALYSIS IN THE AFRICAN REGION - AN APPRAISAL OF THE QUALITY OF DATA

1. INTRODUCTION

1. In the late fifties and the sixties, most of the countries in the region lacked data and information for carrying out plans and policies of economic and social development. Thus the immediate task was to organize data collection in as much detail as was feasible and as expeditiously as possible. Hence the focus upto the early seventies was to fill up gaps in the quantitative aspects of data. Accordingly, almost all of the countries took a census, carried out surveys, strengthened the administrative machinery for the collection of several types of data like those on education, manpower, health, housing, food etc. and tried to improve the system of vital registration.

2. Today there are only three countries which have not taken a census, but all indications are that by the end of the cycle of the 1980 round of censuses i.e., by 1985 most of them would have completed at least one census - some may have two or more post independence censuses. Side by side with extension of geographic coverage, there has been an expansion in the types and varieties of information collected. Thirteen countries also participated in the World Fertility Survey and the picture of the demographic situation in the region is becoming clearer. Recently interest in household surveys has been shown by many countries and it is hoped that in the near future a large amount of information would become available.

3. As the planning needs multiplied, so did the requirements in terms of data and efforts in the direction of data collection were intensified. For proper planning, policy formulation and implementation of programmes it is obvious that the quality of information should be acceptable. Erroneous or biased data would lead to wrong decisions resulting in waste of resources, time and in many instances culminate in human suffering. Even though it cannot be ensured that field data are 100% accurate, it should be possible to ascertain the range of variation. Efforts should be made to reduce the range to the minimum within the constraints of time and resources.

4. Statisticians in the region have been aware of problems in their data and have generally attempted to give a frank appraisal in their reports on these enquiries. For example, many countries have prepared administrative/methodology reports spelling out the problems faced in the collection, compilation and analysis of the data. The lessons learnt have been useful not only to themselves in their later efforts, but also provided useful guidelines for data producers in other countries. Another salient aspect has been the preparation of analytical/evaluative reports and monographs which have brought out the qualitative aspects of the data and offered suggestions and recommendations for future efforts.

5. Thus side by side with attempts to produce the reports in a language understandable to users, publications have been more extensive and intensive. The diffusion seminars and other channels of communication have aimed to reach large segments of the population.

6. When little was known of the demographic, socio-economic and geographic profile of the population, it was very difficult to assess the quality of the information, but with more information, and improved techniques for analysis, it has been possible to evaluate the results and make some adjustments, corrections and other needed manipulations to the data so that they could be utilised for most of the planning needs, at least at the national level.

7. If the sixties and seventies were devoted for collecting as much detailed information as was necessary or feasible, the eighties should consolidate the position by attempts to improve the quality of the data. For this purpose it is necessary to know the types and varieties of errors, biases and deficiencies in the data and their sources so that remedial actions could be instituted. This paper attempts to bring out some of these as noted from analyses of data from the region.

2. TYPES AND VARIETIES OF DATA COLLECTED

8. According to information available in the files at ECA, Table 1 shows the situation in regard to the types and varieties of data collected during recent censuses in the countries of ECA region. It is easy to note that all the countries attempted to cover the basic information recommended by UN and ECA. Local needs and priorities determined additional or special topics or items. Whereas some countries combined census of population with that of housing, others conducted these two operations separately. A few added items bearing on agriculture and live stock in the population/housing census and there were also full fledged agricultural censuses in the region. Thus, there is evidence to suggest that there has been extensive collection of demographic, socio-economic and geographic data and the data are becoming available to the planner and programme administrators.

9. Along with the increase in the number of items canvassed, there has also been expansion in the scope and details of tables prepared. Computers and their peripheral equipment have played an important role in this and also in the preparation of sample data tapes which have been available to researchers for detailed or specialised studies.

10. Censuses have been supplemented by surveys, administrative operations and in a few cases by vital registration systems. In this connection, the role played by WFS is worth mentioning and it is hoped that more countries would carry out similar surveys not only on fertility but also on mortality and population distribution/migration. Two of the most serious population problems perceived by African governments based

Table 1

Number of countries collecting specified information in their census during the period since 1975 (Based on reports from 38 countries)

A. Demographic characteristics

Characteristic	No. of countries	Characteristics	No. of countries
Sex	All	Children living/dead	22
Age	All	Births past year	24
Relationship	36	Survival of past year birth	23
Marital status	36	Parent survival	10
Age at marriage/ duration of marriage	1	Survival of spouse	1
No. of spouses	3	deaths in h.h.	13
Children ever born	25	Physical status/ disability	2

B. Geographic/migration characteristics

Characteristics	No. of countries	Characteristics	No. of countries
Usual place of residence	28	duration of residence	16
Place of birth	34	place of origin	5
Place of previous residence	10	place of work	4
Residence at specified time	14	residence status - present/absent	21

C. Social/personal characteristics

Characteristic	No. of countries	Characteristic	No. of countries
Citizenship/ nationality	31	Language/spoken	31
Religion	10	School attendance	32
Ethnicity/tribe	14	Educ. attainment	32
Literacy	21	Educ. qualification	10

Table 1 (cont'd)

D. Economic characteristics

Characteristic	No. of countries	Characteristic	No. of countries
Activity status economic activity	24	Empl. status	29
Occupation	31	Sector of empl./type of establishment	5
Industry	27	Secondary occ.	4

E. Housing characteristics

Characteristic	No. of countries	Characteristic	No. of countries
No. of house/dwelling units	6	Tenure/type of occupancy	18
Type of housing unit	13	Source of water	20
Quality of housing/ condition of dwelling unit	2	Toilet facility	17
Construction material of wall	20	Bathing facility	4
Construction material of roof	19	Cooking facility (kitchen)	7
Construction material of floor	11	Source of fuel	5
Age of building	6	Source/type of lighting	16
No. of rooms	12	Transportation	6

F. Community level variables

Variable	No. of countries	Variable	No. of countries
Source of water supply	2	Availability of primary school	5
Availability of electricity	2	Distance from nearest hospital/ clinic/health centre	1

on the third and fourth enquiries by UN on the population situation were health/mortality and uneven population distribution/migration.

11. Censuses, surveys and administrative operations cannot supplant the important role played by vital registration system. Every effort must therefore be made to assist governments to improve the system.

12. Data evaluation, analyses and utilisation have, however, not kept pace with data production. To give maximum benefits to countries in the region, it is imperative that concerted effforts be made in the 1980s to encourage fuller analyses and utilisation of the available information. Training of personnel would be a step in the right direction. Analytical monographs illustrating data evaluation, analyses and utilisation also would be helpful.

3. QUALITY OF DATA

13. One of the fundamental information any country needs is the size of its population. Even though it is impossible to obtain a precise figure, most countries would be interested to have information on its population size as on a particular date, perhaps to the nearest hundreds or even thousands. Since basic demographic characteristics like age and sex influence and are influenced by socio-economic factors, it would be desirable to have the age-sex distribution in as much detail as possible to serve various needs like planning for education, housing, health, manpower, food supply and so on. An important aspect of population is its dynamics, i.e., its change over time and hence information on components of population change, viz., fertility, mortality and migration are essential for estimating future population size, composition and distribution. Characteristics like education, economic activity, household headship etc. are very necessary forplanning socio-economic programmes. Thus for proper planning and programming it is essential to have ideas about the accuracy or reliability of information on these various aspects of a population. The experience in the region is illustrated in this paper by a few examples. The intention is to bring out the important aspects of the situation so that the lessons learnt could benefit future producers and users of similar data.

3.1. Population Count

14. With only fragmentary information available when the first post independent national censuses were undertaken in some countries, the population figures looked suspicious and resurveys, analytical techniques etc. had to be applied to ascertain the reliability of the reported figures. For example, in the 1973 census of Sudan because of very low growth rates, underenumeration was suspected in respect of the nomadic population, of the population in Blue Nile and in the Southern provinces. At the same time, analyses and information indicated that some of the foreign born reported themselves as Sudanese in Kassala province. In Botswana, the very high growth rate of population between 1971-1981 has been explained by better enumeration in 1981, improved health conditions resulting in drastic fall in death rate with no consequent fall in fertility, inclusion of refugees and foreign population, return migration etc. It is well known that in Nigeria the figures from the 1973 census were discarded because of the possibility of sizeable errors similar to those found in the 1963 enumeration.

15. Coverage evaluation is very difficult especially when independent sources are either not available or are equally or more biased. To ensure a good enumeration certain basic elements like cartographic preparation, publicity, special enumeration techniques for groups like nomads and remote area dwellers, training of supervisors and enumerators etc. should be stressed. Political and other considerations may bias census returns and hence should be avoided. For example, omission of questions on ethnicity, religion, language etc. may, in certain countries, assist in procuring better quality data.

16. Post enumeration survey is an important tool in the coverage evaluation of a count but it has been used only in very few instances with some success. For example, Liberia (1974) and Cameroon (1976) are among those who claim to have been able to assess the coverage of their population by sex and broad age segments. In the case of Malawi (1966) and Kenya (1969), the PES clearly indicated underenumeration in the census and these formed the basis for correcting the census totals. In Ghana the 1960 census was followed by a PES combining coverage and content evaluation and collection of additional information. The initial effort was not very successful in respect of coverage evaluation but a detailed reconciliation check on 1/2 percent of the sample E.A's showed an over-enumeration of between 0.6 and 2.5 percent. The 1970 coverage error check did not yield any better results due to problems of matching arising out of lack of uniqueness of names. In the 1980 census of Zambia and the 1981 census of Botswana there were PES but from preliminary analyses it is obvious that the main objective of coverage evaluation may not be attainable. Since most countries lack other independent source for checking accuracy of count, PES may be utilised for this purpose as it is the only direct method of coverage evaluation. But it is to be stressed that it should be carried out with due care so that it serves the purpose. Indirect methods such as balancing equation and comparison with external sources may serve only supplementary roles and should not be utilised as a sole aid for checking coverage of an enumeration because of possible errors in the other sources.

17. Coverage evaluation may also be undertaken through techniques of demographic analysis. For example, coverage of certain segments of a population like young children, young adults, females etc. could be detected through analyses of age-sex distributions. Unfortunately, most of the evaluation in the region has been through consistency-convergency techniques and in situations where very little is known of the range of variation within which the true value should lie, the application of such techniques becomes a matter of judgement.

18. As in other parts of the world, enumeration completeness has been selective by age, sex and geographic distribution of the population. Very young children, young adults, females (or males) and mobile groups or remote area dwellers have been more omitted than others. With the available information, it has not always been possible to arrive at absolute figures of omission or duplication but some approximate range could be estimated because of availability of a larger volume and time series of data and more sophisticated analytical tools.

19. There has also been observed to be selectivity in respect of characteristics of the individuals omitted or included. For example, enumerators have been aware of persons in single member households, of young unmarried females, infants and children in high mortality areas, nomads and remote area dwellers. In one census, omission was because a rather long questionnaire covering additional topics was used. Another reason for omission has been the number of rows in the questionnaire which has been noted to result in additional members being not recorded. Similarly, omission of special groups due to benefits/disbenefits also is not unknown. Instances of double counting of persons have been noted due to the de jure method of enumeration especially in joint and extended family systems. Omission/duplication may be consequent on inclusion/exclusion criteria in counting. For instance, if rules are not complementary then there could be duplication or omission depending on the relative stringency of inclusion/exclusion criteria. Exaggeration numbers has been noted because of anticipated social, economic or political benefits. There have been instances of aliens reporting themselves as natives or hiding from enumerators and of people migrating to be included or excluded from counts. Sometimes, natural phenomenon like drought, floods etc. may disturb population settlement and affect count.

20. Thus there are several reasons for an enumeration to be affected and every effort must be made to ensure coverage completeness. As prevention is better than cure, it is advisable to keep these possibilities in mind and initiate remedial actions instead of waiting to complete the count and adjust the data. If the correctness of a population count is suspected, then the figure is of less value for characteristics and distributional aspects of the population as it would be difficult to ascertain which segment is affected. The main focus of a census should be to obtain as accurate a figure as is possible and then only comes the question of composition, distribution, dynamics etc.

3.2. Age and Sex

21. Two of the most important basic demographic variables are age and sex composition of a population. Not only do they determine the dynamics of the population, they also are determinants of several social, economic, cultural, geographic and physical aspects of a population. Hence accuracy of information on age and sex is fundamental and caution should be exercised in the collection, compilation, interpretation and utilization.

22. Age is one of the most important variables for which accurate information is difficult to obtain in most of the countries in the region. This has been due to less importance given in the societies to an exact age rather than to a gradation of cohorts. Since most countries lack continuous registration of vital events and have only hazy ideas of the level, pattern and trends of vital parameters, a reasonably accurate age data would have been of immense value for the estimation of vital parameters. Since it may be easier to improve age data, efforts should be put in this direction side by side with strengthening the vital statistics system.

23. There should not be any difficulty in the collection of information on the sex of an individual. However, the actual situation is different. Several factors like cultural patterns, indifference, deliberate misstatements or omissions intervene to bring in

deficiencies in sex reporting of a population. Recording and processing errors also have been noted. Since sex ratio at birth and of a closed population are quite stable and not difficult to estimate, it is possible to assess errors, biases etc. in data and adjust accordingly. However, it is always preferable to have reasonably acceptable data from the field itself rather than to adjust an erroneous one.

24. As mentioned in connection with coverage completeness, there are certain age-sex segments which are less completely enumerated than others. For example, it has been a more or less general observation that infants and very young children are omitted from the count. It has even been noted that in a few instances, the number of births reported in the households during the past year taking account of deaths reported among these are quite different from the number of infants reported on the schedule. Since in most cases birth-death reporting may be deficient and usually on the lower side, a reported number of infants less than the survivors of births of past year is a clear indication of omission or age reporting errors. Even at aged 1-4 there have been deficiencies. Either it could be outright omission, age shifting or duplication. The errors also have been more for one sex or the other. In the 1978 census of Tanzania, for example, there were reported to be 804317 births in the past year as against an infant population of 697690 implying an implausibly high infant mortality rate of 333. On the other hand in Botswana 1981, there were only 39173 births but 42702 infants were reported. Whereas sex ratio of infants was much above 100 in Libya and Sudan from their 1973 censuses, it was less than 100 in Tanzania 1978 and it was near 100 in Botswana 1981 and Kenya 1979. One of three possibilities exist in cases where the sex ratio is abnormally high or low (i) omission/duplication of infant of one sex (ii) age shifting and (iii) wrong reporting of sex. In de jure counting especially in societies where young children spend a long time in grand parents houses, it is possible for these children to be reported in more than one household. Duplication also is possible if some benefits like free milk, medicines etc. are given. Due to misunderstanding, misinterpretation or incomplete recording, cases have been met where a child of under one year is reported as one or more years old. For example, a child may be reported as 7 months old and recorded as 7 and coded and tabulated as 7 years. Instances of many children aged less than one year whose actual month of birth or number of months since birth were not known, were reported as 11 months old and tabulated as 11 years, depleting the infant population and bloating up those in age 11. Again it has been observed that respondents are unable to understand how a child can be 0 years old and hence report infants as 1 year old. All these could be sex selective. In some cases the sex column may be left blank because the name suggests the sex and at the processing stage errors could crop in. There are several names similar between the sexes and sometimes the difference may be only in the last vowel, as in Arabic or Hispanic names. Hence if writing of names is not clear, interchange of sex becomes a possibility.

25. Thus some infants are omitted outright and some are shifted to another sex group or age group or both. It has been found that even the children aged 1-4 years are much less than what one would expect from the fertility-mortality conditions. In addition to omission, there could be shifting to the higher age. Since infant and child mortality, fertility levels and sex-ratio at birth are not accurately known, it is imperative that every avenue to ensure a good count of the very young ages should be sought.

26. Some of the omitted child population is reported in age 5-9 years and in a large number of instances children aged 10 and above also are reported as aged 5-9 years because of age determination for education and unwillingness to fill the long questionnaires which are usually for those above age 10. This phenomenon thus depletes those aged 10 and above unless partially compensated by wrong transfer of children 10 and 11 months into this age group. For instance, based on the population aged 0-4 the 1973 census of Libya indicated a birth rate of 52.8 but the 5-9 age group showed a birth rate of 59.2. In Lesotho the corresponding values were 34.2 and 37.8. In Libya in addition to the above factors indicated for age shifting, there was an additional reason based on the census interval. Ages were estimated in most cases based on an identity document issued in 1964 and in 1973 digit 9 became preferred. Similar digit preference results from asking date of birth and use of event calendars.

27. In the young adult ages omissions occur because of mobility, marital status, and other socio-cultural reasons. For instance, a female who is married may be reported as older and one who is still single may be shifted to a younger age. Mobility may result in some members of households not being reported as is suspected in the 1967 census of Tanzania. In the earlier censuses of Libya, omission of females was noted especially in young adult ages. In the reproductive age group, males are shifted to older ages and females are shifted from the younger and older ages into the reproductive ages with the result that in most countries the sex ratio in this age segment, i.e. 15-49 years is much lower than 100.

28. At older ages the tendency is for age exaggeration and this is observed to be more for males than for females. This results in abnormally high sex ratios at ages 70 and above and also the proportions at these older ages look incompatible with estimated mortality levels.

29. The net effect of these shifts are seen in fluctuating age and sex ratios culminating in a large joint score, a rather high mean age of the population, high mean age at birth of children and of fertility, high survival ratios at older ages and excessive values of total fertility rates.

30. Table 2 presents some indices of age-sex data quality. It is clear that only one country has a UN joint score less than 20 whereas 2 others have values very near 20, i.e. the data can be considered as accurate. Around two thirds of the countries, for which age-sex data were available, indicated 'highly inaccurate' data.

31. Another index (the Myer's digit preference index) also indicated similar results. For example, the rank correlation between the UN joint score index and the Myer's indices were near .8 for both sexes. This value of the correlation coefficient is highly significant.

32. Table 3 presents these indices from their 2 recent censuses. A look at the values clearly show that there has been improvement in single year and five year age-sex data.

Table 2. Age ratio, sex ratio and Joint Scores, Myers indices and data quality rating

Country	Census Year	Age Ratio Score M	Age Ratio Score F	Sex Ratio Score	Joint Score	Quality Rating	Myers Index M	Myers Index F
Algeria	1977	5.3	5.5	3.1	20.1	B		
Egypt	1976	10.2	14.2	8.1	48.7	C		
Morocco	1971	14.2	25.2	16.2	88.0	C		
Socialist People's Libyan Arab Jamaharia	1973	5.6	11.0	6.1	34.9	B	9.8	25.8
Sudan	1973	17.7	20.7	9.6	67.2	C	48.2	55.0
Tunisia	1975	6.9	5.2	5.3	28.0	B	.	.
Cape Verde	1970	12.7	12.6	5.8	42.7	C		
Gambia	1973	15.1	25.3	14.7	84.5	C	44.7	51.0
Ghana	1970	8.3	11.6	7.6	42.7	C	26.8	31.2
Ivory Coast	1975	5.8	9.2	10.6	46.6	C		
Liberia	1974	11.2	14.6	11.8	61.2	C	27.5	29.4
Mali	1976	9.6	19.5	12.2	65.7	C		
Nigeria	1963	29.0	36.4	14.1	107.7	C		
Senegal	1976	4.5	5.6	8.4	35.3	B		
Sierra Leone	1974	13.0	16.1	8.5	54.6	C	44.4	47.2
Togo	1970	18.7	19.0	11.6	72.5	C		
Upper Volta	1975	10.1	14.1	10.1	54.5	C		
Cameroon Un. Rep.	1976	5.6	6.8	5.6	29.2	B		
Central Afr. Rep.	1975	18.9	26.2	9.7	74.2	C		
Congo	1974	4.9	4.1	5.3	24.9	B		
Gabon	1960	11.4	13.2	9.9	54.3	C		
Botswana	1981	3.4	5.2	3.6	19.4	A	7.1	6.8
Djibouti	1983	15.8	20.5	9.9	66.0	C		

Table 2 (Cont'd)

Age ratio, sex ratio and Joint Scores, Myers indices and data quality rating*

Country	Census Year	Age Ratio Score M	Age Ratio Score F	Sex Ratio Score	Joint Score	Quality Rating	Myers Index M	Myers Index F
Kenya	1979	3.6	4.0	4.2	20.2	B	12.4	15.4
Lesotho	1976	9.2	8.7	5.3	33.8	B	10.4	9.6
Malawi	1977	14.4	9.2	8.3	48.5	C	12.9	16.6
Mauritius	1972	8.7	8.4	2.9	25.8	B		
Mozambique	1970	10.9	13.6	8.0	48.5	C		
Seychelles	1977	4.4	7.8	7.6	35.0q	B		
Somalia	1975	43.6	50.9	14.7	138.6	C	61.6	68.6
Swaziland	1976	8.9	5.2	5.6	31.9	B	14.1	14.7
Tanzania Un.Rep.	1978	10.0	9.0	3.8	45.4	C	25.2	32.1
Uganda	1969	8.3	12.3	8.9	47.3	C	24.6	33.3
Zambia	1969	14.4	8.0	14.6	66.2	C	14.1	15.3
Zimbabwe	1969	10.0	9.9	3.8	46.3	C		

The quality rating is: A: Score between 0 and 20 - Accurate
B: Score between 20 and 40 - Inaccurate
C: Score above 40 - Highly inaccurate

Table 3. Age ratio, Sex ratio and Joint scores for selected countries from their recent two censuses

Country	Year	Age ratio Score M	Age ratio Score F	Sex ratio Score	Joint Score	Myers Index M	Myers Index F
Ghana	1960	12.4	14.4	8.8	53.2	29.6	31.2
	1970	8.3	11.6	7.6	42.7	26.8	31.2
Kenya	1969	5.1	7.4	6.7	32.6	21.2	22.2
	1979	3.6	4.0	4.2	20.2	12.4	15.4
Lesotho	1966	10.0	9.7	6.5	39.2	11.8	12.8
	1976	9.2	8.7	5.3	33.8	10.4	9.6
Libya	1964	13.4	21.1	14.4	77.7	46.5	70.3
	1973	5.6	11.0	6.1	34.9	9.8	25.8

33. Usually the enumerator has to estimate the age of persons in the household (e.g. at the 1974 census of Liberia it is stated that more than 90% of ages were estimated by enumerators) those whom he sees and many he does not see, using a calendar of events, age-grade cohorts, physiological changes, documentary or other evidence, physical appearance, marital status, fertility level, hierarchical position in household, relationship of members and so on. Inspite of all these, many persons may be reported as 'age not stated'. Imputation and other tools of adjustment are risky and can vitiate any findings. Especially if the omission is affecting as large as 17% as noted in the 1974 sample census of Zambia, the age data becomes of very little value. It is easier and much safer to estimate age in the field itself, however approximate it may be. Another important matter worthy of mention here is that inspite of the fact that single year of age data would be subject to large magnitude of errors, biases and deficiencies, it is advisable to collect it and present it. This would enable the analyst to arrive at a better picture of the true situation in the country than if he were only provided with quinary or other broad age group data. Appropriate techniques exist for evaluating and adjusting such data.

3.3. Relationship, Marital Status, Fertility, Mortality and Physical Status

34. The question on relationship inquired in censuses is not only for checking consistency of data or for ensuring completeness of count, it has a very important socio-cultural utility on its own. Hence efforts must be made to make the information available by proper coding of the information, tabulating it in required details etc. For example in the 1976 census of Lesotho information on relationship of members was collected, but at the coding stage it was left out, thus depriving the country of an important piece of sociological information. In others, tabulations are not prepared in sufficient detail to draw the maximum benefit.

35. In addition to conceptual problems in the definition of marital status, there are problems in the reporting and recording of the information. There is a tendency for some divorced or separated persons to be reported as 'single' or 'never married'. Then there are problems brought in by differential coverage by sex and age reporting errors.

36. Fertility data is usually affected by underreporting and age misstatements. In several instances confusion between 'zero parity', 'not stated' and 'not applicable' have affected the information. An instance where only female children ever born were recorded because the question was for females only has been noted and in another instance grown up children were omitted because 'they were no more children'. Coding has had its due share in vitiating information on parity. For instance in the 1976 Lesotho census, the 'not stated' and 'not applicable' cases were given the code 99 which resulted in serious problems in parity data. Similarly, in the 1978 Tanzania census the same code for 'not applicable' and for those with '8 or more children' has been noticed to have created analytical problems. It is advisable to tabulate parity in such details as to derive the total number of children, if the figure is not explicitly given in the tables especially when a large percentage of women are reported with high parity. Coding, punching and other data processing errors also have resulted in high parities at young ages. Omission of children living elsewhere and those dead is also not uncommon. It is suggested

that children ever born should be collected and tabulated by sex of children and details like those living here, those living elsewhere and those dead should be included in the questionnaire for consistency and other checks. Memory lapse is not only for past events, it can occur even for recent events. As noted earlier in the case of Botswana the number of births of a recent year could be less than the reported number of infants. Even the latter may be an under reported figure. There is also the problem of reference period error and omission of children who died in early infancy. One indication of memory lapse can be noticed from the number of births reported for past years in the 1973 census of Sudan where the ratio of those reported for the period two, three years before census to those reported for the current year was only 0.72 and 0.60 respectively.

37. The question on deaths in the household in the past year has not been successful in most cases. Thus mortality information has to be based on indirect methods like those developed by Brass et al on child, spouse, sibling and parent survival.

38. Very few countries canvassed information on physical status, disability etc. Inspite of ECA finding that 'questions on disability (mental or physical) can best be investigated in a health survey with specially trained interviewers', a few countries collected the information and used it for planning purposes.

3.4. Geographic characteristics, Migration, Ethnicity, Nationality, Language Religion etc.

39. The spatial distribution of a population is very important and as indicated by member countries in the region the problems arising from uneven population distribution have been perceived to be major. Apart from enumeration completeness by geographic aspects, there are problems in analysis and interpretation of data due to changes of place names and boundaries of civil divisions. This is compounded by difficulties in presentation of data at small area levels. Countries may start with ambitious schemes in identifying persons at very small geographic levels, but when it comes to coding, tabulation etc. some of these may have to be compromised. An example of this is from the 1973 census of Sudan where at the final tabulation stage only provincial urban-rural information was presented, even though information was collected in very great detail in the questionnaire. Some countries discontinued collecting information on ethnicity, tribe etc. because of socio-political and other reasons. Admitting that the information has great sociological importance, if it would affect data quality, it is advisable to delete it from census enquiries. The same may be true of religion and language. The question on nationality is yet another one which has created suspicion. In some countries, as for instance in Sudan 1973, nationality may have been wilfully reported wrongly to avoid problems or to get some benefits. In others as suspected in 1970 census of Ghana, omission of some nationality groups may be a great possibility. Place of birth is a simple question to answer, but when someone else reports for members of households, the answers may not be correct. Deliberate misreporting perhaps may be only for foreign born or minority groups. The question on rural-urban nature of place of birth could create confusion as to whether the answer should pertain to present situation or situation at the time the respondent moved.

3.5. Education and Economic activity

40. Reference has already been made to the fact that age data get vitiated because additional information is canvassed only on certain segments of a population. Also mention was made of age estimation being affected by schooling status. While it is clear that the reporting of age-sex characteristics would be better among those who are attending or have attended school than among the others, certain types of errors are peculiar to the former. For instance, an observation has been made that in countries where age or date of births is unknown or is rare, children are assigned dates of birth by schools at admission. One would find large numbers of children with same date and months of birth but perhaps with different calendar years. Age shifting could result from such a phenomenon. A problem of those who were in a terminal class previous to a census which was conducted during the holidays is that many would be reported either as in the school or outside depending upon how the question is asked. For females, it has been noticed that it reduces enrollment ratios. The question on literacy is well known to be difficult to obtain accurately because of several reasons. Another problem is the tendency for persons to report the standard attending as that attained or vice versa. Category climbing is noticed in respect of question on professional qualification and highest grade/level completed.

41. The problem in the collection of economic activity data is even more widespread. Age reporting errors could bring in a saw tooth pattern in participation rates and there is also the problem of selective omission by age and sex due to one reason or another. Here also two tendencies have been noted. If the additional work involved in canvassing the long questionnaire containing economic information is not properly remunerated, the tendency is to push persons to lower or higher ages to avoid the extra work. On the other hand, if payment is according to number of questionnaires filled in, then the opposite tendency of pushing young and old persons into the economically active age might occur. Definition of the concept is yet another source of great difficulty to those carrying out data collection - whereas in some countries almost all of the females are reported as economically active, the other extreme of very few being reported also do occur.

4. CONCLUSIONS AND RECOMMENDATIONS

42. Analysis of the data from censuses in the region have indicated that the usual types of errors and deficiencies were prevalent. However, there is indication that the magnitude of these deficiencies have reduced and definite improvement is shown in age-sex reporting. The evaluation and analysis indicated that some of these errors could still further be reduced and efforts should be focussed in this direction.

43. There has not been enough analyses and evaluation of data commensurate with data availability. It is recommended that greater attention be given to a fuller evaluation and analyses of the data collected.

44. Comparative analysis of data from the various countries bringing forward the experiences of diverse countries seem to be worth attempting. The experiences of other countries could be useful in improving the data in countries that have problems of data quality.